

Department of Electrical & Systems Engineering

Departmental Papers (ESE)

University of Pennsylvania

Year 2008

Challenges of Country Modeling with
Databases, Newsfeeds, and Expert
Surveys

Barry G. Silverman^{*}

Gnana K. Bharathy[†]

G. Jiyun Kim[‡]

^{*}University of Pennsylvania, barryg@seas.upenn.edu

[†]University of Pennsylvania

[‡]University of Pennsylvania

Postprint version. Published in *Multi-Agent Systems: Simulation and Applications*, edited by Adelinde M. Uhrmacher and Danny Weyns (Boca Raton, Fla.:CRC), 2008.

This paper is posted at ScholarlyCommons.

<http://repository.upenn.edu/ease-papers/455>

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2008		2. REPORT TYPE		3. DATES COVERED 00-00-2008 to 00-00-2008	
4. TITLE AND SUBTITLE Challenges of Country Modeling with Databases, Newsfeeds, and Expert Surveys				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Pennsylvania, 3451 Walnut Street, Philadelphia, PA, 19104				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Challenges of Country Modeling with Databases, Newsfeeds, and Expert Surveys

Barry G. Silverman
University of Pennsylvania

Gnana K. Bharathy
University of Pennsylvania

G. Jiyun Kim
University of Pennsylvania

9.1	Introduction.....	9-1
9.2	Cognitive Agent Modeling	9-4
	Major PMF Models within Each PMFserv Subsystem	
9.3	Social Agents, Factions, and the FactionSim Testbed	9-8
9.4	Overview of Some Existing Country Databases...	9-12
9.5	Overview of Automated Data Extraction Technology	9-14
9.6	Overview of Subject Matter Expert Studies/Surveys	9-19
9.7	Overview of Integrative Knowledge Engineering Process	9-20
9.8	Concluding Remarks.....	9-26
	Acknowledgement	9-27
	References	9-28

9.1 Introduction

According to expert practitioners and researchers in the field of human behavior modeling ([Silverman et al., 2002; Pew and Mavor, 1998; Ritter et al., 2003]), a common central challenge now confronting designers of HBM (human-behavior-modeling) applications is to increase the realism of the synthetic agents' behavior and coping abilities. It is well accepted in the HBM (human-behavior-modeling) community that cognitively detailed, "thick" models are required to provide realism. These models require that synthetic agents be endowed with cognition and personality, physiology, and emotive components. (We will hereafter refer to these rich models as "cognitively detailed models" or "thick agents.") To make these models work, one must find ways to integrate scientific know-how from many disciplines, and to integrate concepts and insights from hitherto fragmented and partial models from the social sciences, particularly from psychology, cultural studies, and political science. One consequence of this kind of integration of multiple and heterogeneous concepts and models is that we frequently end up with a large feature space of parameters that then need to be filled in with data.

glad

9-493-8-007-0/01/\$6.00+\$1.50
© 2001 by CRC Press, LLC

In recent years, modeling methodologies have been developed that help to construct models, integrate heterogeneous models, elicit knowledge from diverse sources, and also test, verify, and validate models (see [Bharathy, 2006] for example). However, these methodologies have required extensive use of manual labor to develop each model. The development of automatic techniques would significantly improve the efficiency of this process. In this chapter, we will explore how a modeler can navigate, sift, and harvest the vast ocean of data that today can be accessed with a keystroke in certain cases. We will use country data as our theme. By “country data” we mean all possible sources of information that can be used to instantiate agent-based models of complex social systems. Country data are especially important when we integrate several models in order to build a realistic complex social system. This kind of integration of models tends to produce a large feature space of parameters that then needs to be filled in with data. There is no dearth of country data. However, the challenge lies in finding the right data for the right slots.

The state-of-the-art for extracting relevant data is summarized in Figure 9.1, which shows three parallel extraction pathways including (1) webscraping of newsfeeds, (2) extraction of data from country databases, and (3) self-explanatory expert survey forms. These pathways are the focus of this chapter. We examine what they consist of, how they may be utilized, and what issues and challenges arise as we exploit them to rapidly generate agent-based models. This is an exciting time to be working in this field; important breakthroughs seem possible. At the same time, there are also many unsolved issues, and we review these in this chapter as well. In the end, we believe that there is no single best route to obtaining our information of interest, namely the information we need to determine our model parameters. Instead, it seems wisest to fully utilize all three routes and to try to elicit the best possible information through a careful triangulation. Accordingly, Figure 9.1 thus shows how we conceptualize this triangulation.

To understand these issues, one must take an in-depth look at an exemplary agent-based model and consider its data needs in relation to what the three previously mentioned automated extraction pathways can readily produce. As we explore our example, we will ask: In what ways can the modeling and simulation community best marshal the volumes of country data now being made available in databases assembled by social scientists, area studies specialists, and various governmental agencies, and international organizations (databases that track not only the socio-demographics and politico-economic data, but also significant events and the needs/values/preferences/norms of populations of interest)? Currently, these databases—consisting, variously, of expert and mass surveys and opinion polls, conflict and event databases, socio-cultural and politico-economic indicators, human terrain systems [Kipp et al., 2006], automated scraping of newsfeeds and websites, and more—are not all updated frequently enough to capture the most up-to-date information and may not be user-friendly enough to develop a unified database under a common format. Nonetheless, they are collected and maintained by regional and subject-area experts with in-depth local knowledge and wisdom using cutting-edge survey methods and other reliable data collection methodologies. Moreover, the growing interest in the development of various automated data extraction and consolidation techniques (e.g., General Inquirer [Stone et al., 1966], Kansas Event Data System (KEDS) [Schrodt and Gerner], Opinion Analysis System (OASYS) [Cesarano et al., 2006], Profiler+ [Young, 2001], ReadMe [Hopkins et al., 2007], STORY [Fayzullin et al., 2007], to mention just a few) highlights both the possibility and the promise of making these databases more user-friendly and the actual process of data collection more efficient, especially with regard to capturing real-time news feeds from various (web-based) sources around the world.

These automated techniques currently complement and can substitute to a certain extent most standard, labor-intensive data collection efforts. Our chapter provides a cursory

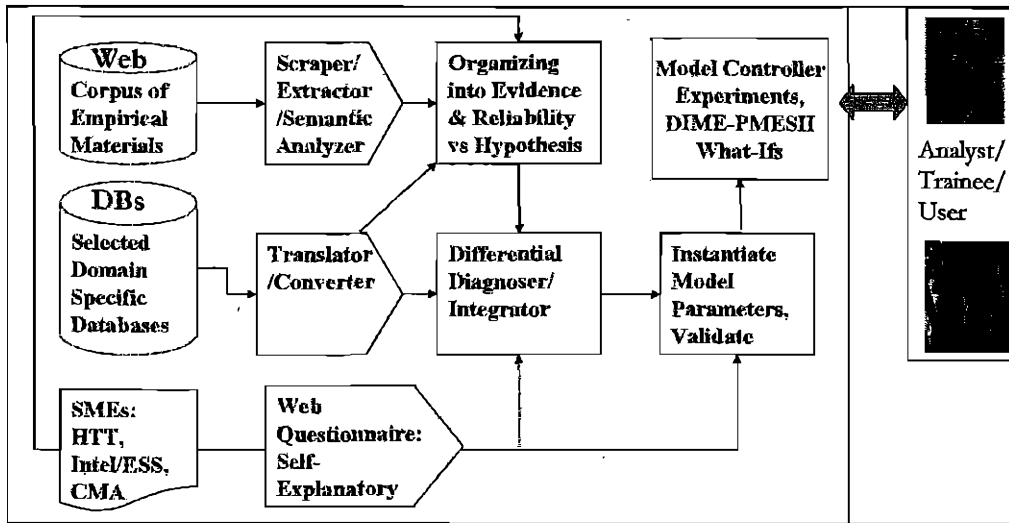


FIGURE 9.1 Overview of New Approaches to Data Extraction and Rapid Generation of Agent-Based Models (Note: SMEs– Subject Matter Experts; DIME– Diplomatic, Informational, Military, and Economic actions; PMESII– Political, Military, Social, Economic, Informational, and Infrastructure effects; HTT– Human Terrain Tool; Intel/ESS– Intelligence collections/Every Soldier a Sensor– a military program; CMA– Civil Military Affairs. Also, all arrows in all figures of this chapter indicate process direction and/or flow of information.)

overview of these databases and techniques and suggests the next steps for the use of this resource. These datasets and techniques are key assets, as we argue, for those interested in the synthesis of two major agent-based modeling paradigms – the cognitive and the social. Consequently, the modeling and simulation community loses a significant opportunity if it fails to tap into this valuable resource. We pursue this argument by means of a case study integrating a cognitive agent environment (PMFserv) and a social agent environment (FactionSim), which we then apply to various countries, regions, and topics of interest (Iraq, Southeast Asia, the Crusades) to assess their validity and realism. Using the information from these databases to populate such models with realistic agents improves their realism and facilitates their refinement. Some information will also be set aside for later empirical testing of these models and their observable implications with a view to achieving external validity

As we explore this new frontier of (auto-generated) agent-based modeling using country databases and newsfeeds, we ask: What can the field of modeling and simulation add to the conventional studies of countries and regions typically performed by social scientists and area specialists? Country databases have been assembled in order to add depth to the study of countries and regions. In order for the data requirements of modeling and simulation to be met by any one of these sources, a dialog must be initiated to determine what sorts of data that the models actually need versus what is now collected. It is also worth studying whether the modeling and simulation community can seamlessly exchange data with various social science communities. Building on our experience at viewing countries as complex social systems, we aim to outline what agent-based simulation might offer. That is, if we use the data from country databases to help model the “parts” and their micro-decision processes, can we observe macro-behaviors emerging that will aid the work of country analysts? We recognize that, if our aim is to model and simulate a social system from the bottom up, then

we need to approach this system with agent technology that covers both the social processes that influence people and also the cognitive processes individuals use as they reason and as they experience emotion. That is, we are interested in discovering what socio-cognitive agents can offer to the study of specific countries or social systems, and we wish particularly to model how diplomatic, intelligence, military, and economic (DIME) actions might affect the political, military, economic, social, informational, and infrastructure (PMESII) systems of a given country of interest.

Finally, as Sun [Sun, 2006] points out in his useful survey of the respective fields of social agents and cognitive agents, there are very few environments that straddle both topics and, consequently, provide socio-cognitive architectures. In this chapter, we illustrate one such architecture to provide insight into its operations, its uses, and the validity of its outputs. More importantly, we argue that this particular socio-cognitive architecture can serve as an ultimate test-bed for evaluating numerous paper-based theories regarding the operations (political, economic, and more) of our countries of interest. We further suggest that all paper-based theories should be tested and implemented in relation to this architecture. While this framework is relatively mature Commercial Off The Shelf (COTS) software, we close with a discussion of future research needs focused on making new software tools better able to support varied analyses of the PMESII (Political, Military, Economic, Social, Informational, and Infrastructure) systems of the country of interest.

This chapter consists of eight additional sections following the introduction and corresponding to many of the blocks of Figure 9.1: Section 9.2– Cognitive Agent Modeling and Major PMF Models; Section 9.3– Social Agents, Factions, and the FactionSim Testbed; Section 9.4– Overview of Some Existing Country Databases; Section 9.5– Overview of Automated Data Extraction Technology; Section 9.6– Overview of Subject Matter Expert Studies/Surveys; Section 9.7– Overview of the Integrative Knowledge Engineering Process (evidence tables, differential diagnosis); and Section 9.8– Concluding Remarks. In its broadest reach, this chapter introduces and explores a new direction for the modeling and simulation community aimed at capitalizing on a potentially rich symbiotic relationship with the social science/area studies community.

9.2 Cognitive Agent Modeling

We will illustrate the data issue using PMFserv, a COTS (Commercial Off The Shelf) human behavior emulator that drives agents in simulated gameworlds. This software was developed over the past eight years at the University of Pennsylvania as an architecture to synthesize many best available models and best practice theories of human behavior modeling. PMFserv agents are unscripted, but use their micro-decision making, as described below, to react to actions as they unfold and to plan out responses.

A performance moderator function (PMF) is a micro-model covering how human performance (e.g., perception, memory, or decision-making) might vary as a function of a single factor (e.g., sleep, temperature, boredom, grievance, and so on). PMFserv synthesizes dozens of best available PMFs within a unifying mind-body framework and thereby offers a family of models where micro-decisions lead to the emergence of macro-behaviors within an individual. None of these PMFs are “home-grown”; instead they are culled from the literature of the behavioral sciences. Users can turn on or off different PMFs to focus on particular aspects of interest. These PMFs are synthesized according to the inter-relationships between the parts and with each subsystem treated as a system in itself.

9.2.1 Major PMF Models within Each PMFserv Subsystem

The unifying architecture in Figure 9.2 shows how different subsystems are connected. For each agent, PMFserv operates what is sometimes known as an observe, orient, decide, and act (OODA) loop. PMFserv runs the agents perception (observe) and then orients the entire physiology and personality/value system to determine levels of fatigues and hunger, injuries and related stressors, grievances, tension buildup, impact of rumors and speech acts, emotions, and various mobilizations and social relationship changes since the last tick of the simulator clock. Once all these modules and their parameters are oriented to the current stimuli/inputs, the upper right module (decision-making/cognition) runs a best response algorithm to try to determine what to do next. The algorithm it runs is determined by its stress and emotional levels. In optimal times, it is in vigilant mode and runs an expected subjective utility algorithm that reinvokes all the other modules to assess what impact each potential next step might have on its internal parameters. When very bored, it tends to lose focus (perception degrades) and it runs a decision algorithm known as unconflicted adherence mode. When highly stressed, it will reach panic mode, its perception basically shuts down and it can only do one of two things: cower in place or drop everything and flee. In order to instantiate or parameterize these modules and models, PMFserv requires that the developer profile individuals in terms of each of the module's parameters (physiology, stress thresholds, value system, social relationships, etc.).

As an illustration of one of the modules in Figure 9.2 and of some of the best-of-breed theories that PMFserv runs, let us consider "cognitive appraisal" (Personality, Culture, Emotion module) – the bottom left module in Figure 9.2. This is where an agent (or person) compares the perceived state of the real world to its value system and appraises which of its values are satisfied or violated. This in turn activates emotional arousals. For the emotion model, we have implemented one as described in [Silverman et al., 2006b]. Implementing a person's value system requires every agent to have its goals, standards, and preference (GSP) trees filled out. Most significant from the perspective of data production are GSP trees. These are multi-attribute value structures where each tree node is weighted with Bayesian importance weights. A Preference Tree represents an agent's long-term desires for world situations and relations (for instance, no weapons of mass destruction, an end to global warming, etc.) that may or may not be achieved within the scope of a scenario. Among our agents, this set of "desires" translates into a weighted hierarchy of territories and constituencies (e.g., no tokens of leader X in resource Y of territory Z).

The Standards Tree defines the methods an agent is willing to employ to attain his/her preferences. The Standard Tree nodes merge several best available personality and culture profiling instruments such as, among others, Hermann traits governing personal and cultural norms [Hermann, a], standards from the GLOBE study [House et al., 2004], top-level guidelines related to Economic and Military Doctrine, and sensitivity to life (humanitarianism). Personal, cultural, and social conventions render inappropriate the purely Machiavellian action choices ("One shouldn't destroy a weak ally simply because they are currently useless"). It is within these sets of guidelines that many of the pitfalls associated with shortsighted Artificial Intelligence (AI) can be sidestepped. Standards (and preferences) allow for the expression of strategic mindsets.

Finally, the Goal Tree covers short-term needs and motivations that drive progress toward preferences. In the Machiavellian [Machiavelli, 1965, 1988] and Hermann-profiled [Hermann, a] world of leaders, the Goal Tree reduces to the duality of growing/developing versus protecting the resources in one's constituency. Expressing goals in terms of power and vulnerability provides a high-fidelity means of evaluating the short-term consequences of actions. For non-leader agents (or followers), the Goal Tree also includes traits covering

year?

year?

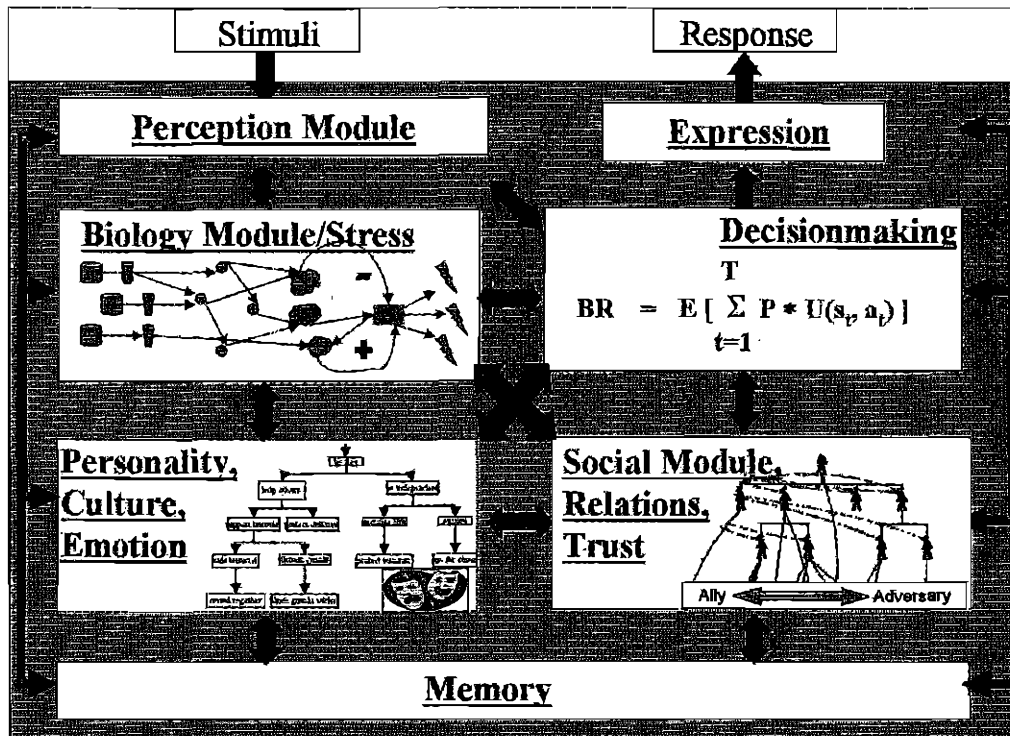


FIGURE 9.2 PMFserv, an example of Cognitive Architecture

basic Maslovian type needs.

Figure 9.3 not only graphically lists some of the example* performance moderator functions (PMFs) in the collection, but also shows how these different functions are synthesized to create the whole (PMFserv). In this sense, Figure 9.3 is simply a more detailed representation of Figure 9.2. The details of these PMFserv models are beyond the scope of this chapter. Interested readers should consult [Silverman et al., 2006b, 2007] for details.

PMFserv has been deployed in a number of applications, gameworlds, and scenarios. A few of these are listed below in Table 9.1.* To facilitate the rapid composition of new casts of characters we have created an Integrated Development Environment (IDE) in which one

*It is worth noting that because our research goal is to study best available performance moderator functions (PMFs), we avoid committing to particular performance moderator functions. Instead, every performance moderator function explored in this research must be readily replaceable. The performance moderator functions that we synthesized are workable defaults that we expect our users will research and improve on as time goes on. From the data and modeling perspective, the consequence of not committing to any single approach or theory is that we have to come up with ways to readily study and then assimilate alternative models that show some benefit for understanding our phenomena of interest. This means that any computer implementation we embrace must support plugin/pluginout/override capabilities, and that specific performance moderator functions as illustrated in Figure 9.3 should be testable and validatable against field data such as the data they were originally derived from.

*Many of these previous applications have movie clips, tech Reports, and validity assessment studies available at <http://www.seas.upenn.edu/~barryg/hbmr>. Several historical correspondence tests indicate that PMFserv mimics decisions of the real actors/population with a correlation of approximately 80% (see [Silverman et al., 2006a, 2008]). Some of these applications are discussed in greater detail in the subsequent section.

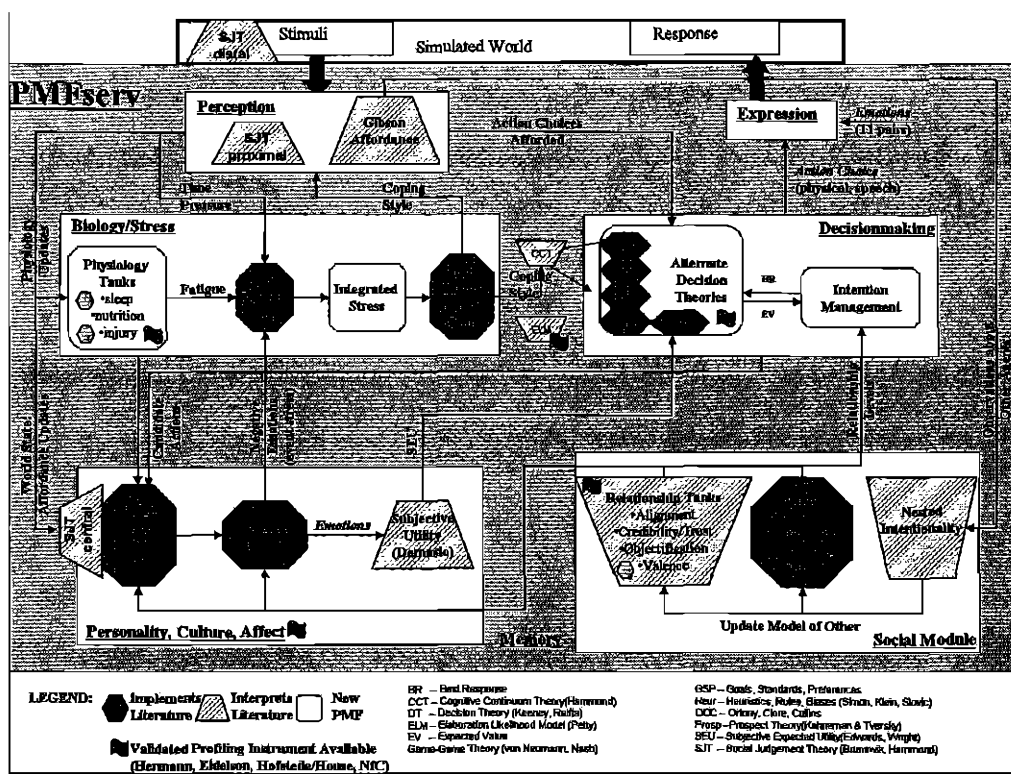


FIGURE 9.3 Summary of Implemented Theories in PMFserv. (Note: NTC– Need for Cognition)

knowledge engineers archetypical individuals (leaders, followers, suicide bombers, financiers, etc.) and assembles them into casts of characters useful for creating or editing scenarios.

TABLE 9.1 PMFserv and Its Applications

Domestic Applications	International Applications
<ul style="list-style-type: none"> • Consumer modeling <ul style="list-style-type: none"> – buyer behavior – ad campaign • Petworld <ul style="list-style-type: none"> – pet behavior • Gang members <ul style="list-style-type: none"> – hooligans • Crowd Scenes <ul style="list-style-type: none"> – milling – protesting – rioting – looting 	<ul style="list-style-type: none"> • Intifadah Recreation (leaders, followers) – Roadmap sim • Somalia Crowds – Black Hawk Down (males, females, trained militia, clan leaders) • Thailand recreation (Buddhists vs. Muslims – radicalization) • Iraq DIME-PMESII sim – three ethnic groups, parliament (leaders and 15,000 followers) • Urban Resolve 2015 – Sim-Red (multiple insurgent cell) • Many world leaders profiled

9.3 Social Agents, Factions, and the FactionSim Testbed

The previous section overviewed the modules of a cognitive agent and some of the components that give it a social orientation. In this section we turn to additional modules that turn the cognitive agent into a socio-cognitive one. Specifically, we introduce FactionSim, an environment that captures a globally recurring socio-cultural “game” that focuses upon inter-group competition for control of resources (Security/Economics/Political Tanks). This environment implements PMFserv within a game theory/PMESII (Political, Military, Economy, Social, Informational, and Infrastructure) Campaign framework. Many of the applications listed above have this game embedded in them. Each group of agents manages the following set of models:

Security Model (Skirmish, Urban Lanchester)

Power-Vulnerability Computations [Johns, 2006]

Skirmish Model (force size, training, etc.)

Urban Lanchester Model (probability of kill)

Economy Model (Harrod-Domar model [Harrod, 1960])

Black Market

Undeclared Market [Lewis, 1954; Schneider and Enste, 2000]

Formal Capital Economy

Political Model (loyalty, membership, mobilization, etc.) [Hirschman, 1970]

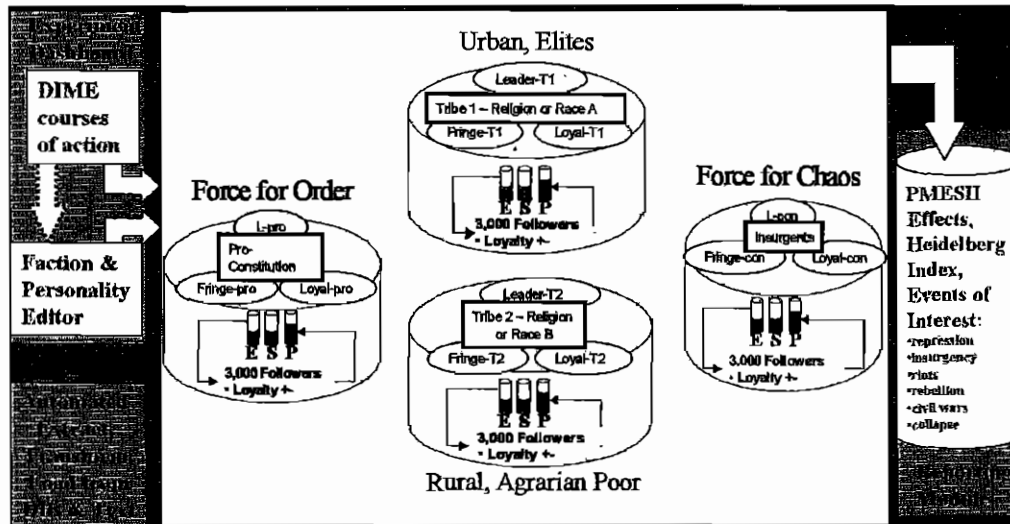


FIGURE 9.4 Models and Components that must be synthesized for a FactionSim Testbed (Note: DIME—Diplomatic, Informational, Military, and Economic; DBs—Databases; Heidelberg Index [Heidelberg Conflict Barometer]; PMESII—Political, Military, Economic, Social, Informational, and Infrastructure)

Institution Sustainment Dynamics

Follower Social Network – Cellular Automata [Axelrod, 1998; Epstein, 2002; Lustick et al., 2004]

Small World Theory/Info Propagation [Milgram, 1967]

This environment facilitates the codification of alternative theories of factional interaction and the evaluation of policy alternatives. FactionSim is a tool that allows conflict scenarios to be established in which the factional leader and follower agents all run autonomously and are free to employ their micro-decision making as the situation requires. A single human player interacts with the environment and attempts to employ a set of DIME (Diplomatic, Informational, Military, and Economic) actions to influence outcomes and PMESII (Political, Military, Economy, Social, Informational, and Infrastructure) effects.

Factions are modeled as in the center of Figure 9.4 where each typically has a leader, two sub-faction leaders (loyal and fringe), a set of starting resources (Economy, E, Security, S, and Political support, P), and a representative set of over 1,000 follower agents. A leader is assumed to manage his faction's E- and S- tanks so as to appeal to his followers and to each of the other tribes or factions he wants in his alliance. Each of the leaders of those factions, however, will similarly manage their own E and S assets in trying to keep their sub-factions and memberships happy. Followers determine the level of the P-tank by voting their membership level. A high P-tank means that there are more members to recruit for security missions and/or to train and deploy in economic ventures. As a result, leaders often find it difficult to move to alignments and positions that diverge very far from the motivations of their memberships.

FactionSim allows one to edit the profiles of all the factions of interest to match a given scenario including:

Faction = { Properties {name, identity repertoire, demographics, salience-entry, salience-exit, other} }

Alignments {alignment-matrix, relationship valence and strength, dynamic alliances}

Roles {leader, sub-leader, loyal-follower, fringe-follower, population-member}

Resources (R) = Set of all resources, r : {econ-tank, security-tank, political support-tank}

r_r, f = {Resource level for resource r owned by faction f , r_r, f ranges from 1 to 100}

$\Delta_{r(a,b)}$ = {Change in r on group a by group b } = Δ_r

T = Time horizon for storing previous tank values

Dev-Level = {Maturity of a resource where 1=corrupt/dysfunctional, 3=neutral, 5=capable/effective}

Actions (A) = { Leader-actions (target) = {Speak (seek-blessing, seek-merge, mediate, brag, threaten), Act (attack-security, attack-economy, invest-own-faction, invest-ally-faction, defend-economy, defend-security)} }

Follower-actions (target) = {Go on Attacks for, Support (econ), Vote for, Join Faction, Agree with, Remain-Neutral, Disagree with, Vote against, Join Opposition Faction, Oppose with Non-Violence (Voice), Rebel-against/Fight for Opposition, Exit Faction}

Despite efforts at simplicity, stochastic simulation models for domains of this sort rapidly become complex. The strategy space for each leader facing only two other leaders grows impossibly large to explore. As a result, FactionSim's Experiment Dashboard (left side of Figure 9.4) permits inputs ranging from one course of action to a set of parameter experiments the player is curious about. On the bottom left is the profile editor governing the personalities of the leaders and sub-leaders, and of the key parameters that define the starting conditions of each of the factions and sub-factions. Certain actions by the player that are thought to alter the starting attitudes or behavior of the factions can flow between these two components, e.g., a discussion beforehand that might alter the attitudes of certain key leaders (Note: this action is often attempted in settings with real Subject Matter Experts, or SMEs, and diplomats playing our various games).

All data from PMFserv and the socio-cultural game is captured into log files. At present we are developing an after-action report summary module, as well as analytical capabilities for design of experiments, for repeated Monte Carlo trials, and for outcome pattern recognition and strategy assessment.

Now, with this framework in mind, let us look at different types of actors required to construct the kind of social system models we have built. Frequently, we create two different types of individual actors:

- individually named personae, such as leaders, who could be profiled, and
- archetypical members* of the society or of a particular group whose model parameters are dependent on societal level estimates.

In addition, we also have groups (collections of agents with leaders and followers) that display some emergent properties of their own that are more than the sum of their parts.

*For each archetype, what's interesting is not strictly the mean behavior pattern, but what emerges from the collective. To understand that, one expects to instantiate many instances of each archetype where each agent instance is a perturbation of the parameters of the set of PMFs whose mean values codify the archetypical class of agent they are drawn from. This means that any computerization of PMFs should support stochastic experimentation of behavior possibilities. It also means that individual differences, even within instances of an archetype, will be explicitly accounted for.

We also model institutions and resources including institutional infrastructures and support plug-in of more detailed models of these dimensions. Typical institutions include the economy (markets, jobs, banking), educational system, the health system, the judicial system, the police and security forces, the utilities/infrastructure (e.g., energy sector, the transportation system, and communication systems), as well as various institutions of civil society.

Types of parameters for typical social system models in PMFserv entities are given below:

Agents (Decision Making Individual Actors):

- Value System/ GSP Tree: Hierarchically organized values such as short term goals, long term preferences and likes, and standards of behavior including sacred values and cultural norms
- Ethno-Linguistic-Religious-Economic/Professional Identities
- Level of Education
- Level of Health
- Level of Wealth
- Savings Rate
- Contribution Rate
- Extent of Authority over the Group

Groups:

- Philosophy
- Leadership
- Relationship to other groups
- Barriers to exit and entry
- Group Level Resources such as Political, Economic and Security Strengths
- Institutional infrastructures owned by the group
- Access to institutional benefits for the group members (Level Available to Group)
- Fiscal, Monetary and Consumption Philosophy
- Disparity

Institutions:

- Capital Investment
- Damage / Decay
- Level of Corruption (indicates usage vs. misuse)

A toolset such as FactionSim (and PMFserv) will only be useful to the extent that it can offer valid recreations of the actual leaders, followers, and populations of interest. In terms of the validity of the current socio-cognitive agent synthesis, this research has tried hard to examine its robustness and cross-sample fitness. FactionSim agents passed validity assessment tests in both of two conflict scenarios attempted to date, as described fully in [Silverman et al., 2008]. In the first scenario, a group of 21 named Iraqi leader agents in 5 factions (FactionSim agents) passed a Turing Test after extensive subject matter expert evaluation by US military personnel, and in the second a separatist movement recreation

involving a SE Asian leader (Buddhist) and his Muslim followers (also, FactionSim agents) passed separate correspondence tests (with correlations of over 79% to real world counterparts). In the version of the Turing test we employed, a group of domain experts attempted to distinguish between the behaviors generated using the models of the agents from those in fact generated by the corresponding actual actors. Consequently, this validation procedure may count as both a rigorous face validation test as well as a Turing test. Validity is a difficult goal to achieve, and one can always devise new tests. A strong test, however, is the out-of-sample test that these agents also passed. Thus the SE Asian leader and his followers were trained on different data than they were tested against. Further, a complete model of leader behavior was originally derived from earlier studies of the ancient Crusades [Silverman et al., 2005] and this model was applied to and evolved into the SE Asian and Iraqi domains. The only features updated were the values of the weights for the value trees and various other group relations and membership parameters – all derived from open sources. So the structure of the leader model also survived scrutiny and passed two out-of-sample tests relative to the Crusades dataset. While these may not be definitive tests, they are sufficient for our purposes at this point as we establish that our descriptive agents are useful components for computational what-if experiments, for training worlds, and to drive agents in third party simulators.

In the subsequent three sections, elaborating on what was presented schematically in Figure 9.1, we will overview the three main sources of empirical information we rely on when building complex social systems using our socio-cognitive agent-based model. The three main sources are: 1) country databases in Section 9.4, 2) empirical materials from the world wide web in Section 9.5, and 3) subject matter experts in Section 9.6. In Section 9.5, we will focus on surveying the kinds of automated data extraction technologies that are available today to obtain empirical materials from the web.

9.4 Overview of Some Existing Country Databases

Existing country databases, broadly speaking, fall into one of two categories.* The first consists of event databases that record significant events of interest in numerous countries around the world. These event databases are valuable resources in terms of providing information about parties and factions, their relative alignment, and the resources on which they can draw in various internal conflict and crisis situations that include civil wars, coup d'états, crackdowns, democratic and non-democratic extrications and internal power transitions, mass killings, terrorist activities, and revolutions. The most up-to-date and extensive event database (in terms of the scope and the extent of database coverage) arguably is the Uppsala Conflict Database (UCD) [Uppsala Conflict Data Program], an expanded version of its predecessor, the Correlates of War (COW) [Sarkees, 2000] database. Both UCD (Uppsala Conflict Database) and COW (Correlates of War) contain both inter- and also intra-state conflict information. Given our present goal of studying complex social systems at the country-level, the discussion will focus on the intra-state event databases. COW (Correlates of War)'s intra-state war data has been known as "the granddaddy of all intra-state conflict datasets" and identifies intra-state wars and their participants between 1816

*Of course, this is only one of the many possible ways to categorize existing country databases. For example, see [Cioffi-Revilla and O'Brien, 2007] for another possible taxonomy of existing databases and, more generally, a good overview of the use of computational analysis in defense and foreign policy research.

and 1997. The UCD (Uppsala Conflict Database) database significantly improves its coverage in comparison with COW (Correlates of War) by lowering the threshold for conflict identification (from COW (Correlates of War)'s 100 annual conflict deaths to 25) and by more frequently updating its database to cover current developments around the world. Any researcher wanting to gather good snapshots of the histories of significant events of interest from around the world should be able to do so with the combined use of COW (Correlates of War) and UCD (Uppsala Conflict Database). At a minimum, these two databases provide some necessary information on relevant faction identification, relative alignment, and some relative resource estimates. Additional information from other excellent event databases such as the Political Instability Task Force (PIT) [Esty et al., 1998], Minorities at Risk (MAR) [Minorities at Risk Project], Atrocities Event Data (AED) [Political Instability Task Force], Intra-state conflict and Interventions (ICI) [Regan, 2000], and Global Terrorism databases (GTD) [Lafree and Dugan, 2007] can certainly supplement and improve the quality of information for events of interest at the intra-state level. In addition to these significant events of interest, some databases such as Protest and Coercion (PCD) [University of Kansas] and Ethnic Conflict and Civil Life (ECC) [Varshney and Wilkinson, 2006] databases even record events of smaller scale (or involving less violence) such as acts of civil disobediences, demonstrations, rallies, riots, sit-ins, work-stoppages, and strikes. Despite the respectable quality and availability of the aforementioned event databases, these event databases by themselves do not provide sufficient information to populate models such as our joint PMFserv/FactionSim mainly because the unit of analysis used by these databases is events rather than factions. The utility of these event databases is thus limited in terms of providing us with quantified snapshots of our events of interest with readily available faction identification, alignment, and strength information, and we need additional information from a second type of database that focuses on country specific opinion polls and mass attitude surveys.

Populating our joint PMFserv/FactionSim framework with realistic agents requires comprehensive and reliable socio-cognitive information about the people of a particular country at the level of our factions of interest. Gathering sufficient information for one PMFserv subsystem of our Value Systems Module, namely, GSP trees, requires a high level of detailed information about people's goals, standards, and preferences, and obtaining this information can be a daunting task. To our relief, we have access to an extensive collection of survey results compiled by survey researchers around the world. The three main publicly available databases in this field are the World Values Survey (WVS) [European Values Study Group and World Values Survey Association], the Global Barometer Surveys (GBS) [Global Barometer Surveys Program], and the Comparative Study of Electoral Systems (CSE) [Comparative Study of Electoral Systems Secretariat]. Both WVS (World Values Survey) and GBS (Global Barometer Surveys) are surveys that are administered in more than 50 countries around the world to measure public opinion on cultural, social, political, and economic issues. The key difference between the two surveys lies in the fact that WVS (World Values Survey) uses a standardized survey questionnaire, while GBS (Global Barometer Surveys) is administered more frequently. Similar to WVS (World Values Survey) and GBS (Global Barometer Surveys), CSE (Comparative Study of Electoral Systems) also tracks public opinion, except that it is held only in countries where there are periodic and reasonably fair elections and focuses on micro-level information on vote choice, candidate and party evaluations, and other relevant information regarding voters' attitudes and values, in addition to standardized socio-demographic measures and aggregate level information on electoral returns and turnouts. It seems, then, that we should be in a position to extract the information we need from these three surveys. Yet, there are two difficulties we face in using the results from these survey instruments for our purpose. The first difficulty lies in

the fact that it is hard to find a one-to-one correspondence between a survey questionnaire item and a parameter of, say, our GSP tree. This is an obvious and unavoidable difficulty given that survey researchers did not design their surveys with our GSP tree parameters in mind. This difficulty, however, is not insurmountable; with some effort, we can select survey questionnaire items that can serve as proxy measures for our parameters of interest. The second difficulty lies in the fact that the unit of analysis for these public opinion surveys is countries while, for many modelers of complex social systems—such as countries—use a unit of analysis that is smaller than a whole country.* For our joint socio-cognitive PMFserv/FactionSim framework, the appropriate unit of analysis is at the faction level. Again, this difficulty that results from the discordance in the unit of analysis can be overcome simply by cross tabulating and sorting these survey databases according to properties that categorize survey respondents into specific groups that match our interests. The surveys are sufficiently detailed to allow us, for example, to infer information about whether an average supporter of a particular political party has a more or less materialistic vision of life than another average supporter of another political party or a different faction. We may even be able to infer a more “elite-level” information (leader-level information within our PMFserv/FactionSim framework) by cross-tabulating our proxy survey items for a particular parameter of interest with the socio-economic information about the respondents, given that leaders are more likely to have higher educational attainment and income and/or are more likely to spend more time in a particular organizational grouping.

In sum, the existing country databases – both the event and the survey ones – are great assets for those of us in the modeling and simulation community who are committed to using realistic agent types to populate our simulated world. However, as noted previously, using existing databases at this stage of their development requires efforts by the modeling and simulation community to study the structure of the available databases and to take into account their strengths, weaknesses, terminology, and idiosyncrasies. In this regard, we need to be creative in finding proxy measures that can reasonably approximate our parameters of interest, and we would need to be imaginative in restructuring our databases in ways that are conducive to extracting the information we want at the level of analysis we want. Finally, it is important to note that the preceding overview of some existing country databases is not exhaustive. There are obviously more event and survey databases than the ones mentioned, not to mention other specialized economic, social, demographic, human capabilities, and political violence databases. We guide interested readers to the Penn Conflict Database Catalogue [Kim and Bharathy] for a more comprehensive overview of the existing databases that may be of relevance to the modeling and simulation community.

9.5 Overview of Automated Data Extraction Technology

As discussed in the preceding overview of existing country databases, our use of these databases is not as efficient and convenient as we might like it to be; nonetheless, these databases are invaluable to our project of building realistic agents and validating our models. Still, it is important to note that we may not be able to gather all the empirical information we need from these databases alone, and we will at times be forced to collect some information by ourselves. Instead of going through the typically labor-intensive data

*One can make an argument that the unit of analysis can also be individual survey respondents in particular countries. In this case, many modelers of complex social systems – such as countries – would be using a unit of analysis that is larger than an individual of a particular country.

TABLE 9.2 A Summary of Some Existing Country Databases

Atrocities Data (AED)	Event	Event	A database collected by Kansas Event Data System to provide a systematic sample of atrocities occurring around the world.
Correlates of War (COW)	Event	Event	This database is known as "the granddaddy of all conflict datasets" and most other conflict event datasets either extend or improve this pioneering database.
Comparative Study of Electoral Systems (CSE)	Survey	Survey	This database provides individual level information on vote choice, candidate and party evaluations, current and retrospective economic evaluations, etc. as well as aggregate level information on electoral returns and turnouts.
Ethnic Conflict and Civil Life (ECC)	Event	Event	This database provides comprehensive information on all Hindu-Muslim riots reported in the major Indian newspapers in India and Pakistan. The database is expanding to cover other countries in Southeast Asia.
Global Barometer Surveys (GBS)	Survey	Survey	This database is one of the two most comprehensive survey databases of public opinion of people around the world. Administered by regions and decentralized.
Global Terrorism Database (GTD)	Event	Event	This database contains information on both domestic and international terrorist events around the world since 1970. For each event, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and the identity of the perpetrator.
Intrastate Conflict Interventions (ICI)	Event	Event	Unlike some other event databases, this conflict database contains information about external/foreign interveners who play a role in intrastate conflicts in addition to the standard information (time, location, involved actors, etc.) in an events database.
Minorities at Risk (MAR)	Event	Event	This database provides comprehensive information on the status and plight of politically-salient communal/minority groups in countries around the world.
Protest and Coercion Data (PCD)	Event	Event	This database provides comprehensive information on protest and coercion in 32 countries around the world. Information concerning date, day, action type, location, protest groups, targets, and the strength of the protesters are provided. This database is collected using the Kansas Event Data System.
Political Instability Task Force (PIT)	Event	Event	This database contains standard event information on political instability events in most countries around the world. The political instability events include ethnic wars, revolutionary wars, genocides and politicides, and adverse regime changes. Adverse regime changes do not include transitions to more open and democratic forms of governance.
Uppsala Conflict Database (UCD)	Event	Event	This database probably is one of the most comprehensive and widely used event databases by social scientists. It contains information on the location of the conflict, the source of incompatibility, opposition organizations, date and duration, conflict intensity level, the nature of conflict settlement, the quantity and quality of arms, and the duration of peace settlement.
World Values Survey (WVS)	Survey	Survey	This database is the other most comprehensive survey databases of public opinion around the world. Unlike GBS, it has a standardized survey questionnaire but is administered less frequently.

collection process with an army of undergraduate and graduate research assistants, we have a new set of tools and technologies that can streamline our data collection efforts, making them less arduous and more efficient. The most promising approach seems to be the use of various automated semantic analysis tools such as Automap [Carley et al., 2006], General Inquirer, Cultural Simulation Model (CSM) by IndaSea [Park and Fables, 2007], Profiler+, Kansas Event Data System (KEDS), ReadMe, The Resource Description Framework Extractor (T-Rex) [Subrahmanian, 2007], OASYS, and STORY to extract our information of interest from various newsfeeds and web scrapings.

The majority of these automated content analysis tools work according to a similar underlying logic. They contain a list of terms of interest together with their synonyms. They then count the frequency with which these terms and their synonyms appear in an actual text to provide us with usable data. For example, if we want to collect information from various newsfeed and web scripts about a faction leader's propensity to use violence, we would build a list consisting of both generic and also highly specific words pertaining to the use of violence by this leader (words ranging from "killing" to "tire necklace," for example) and let these programs count the frequency of such words in various texts. This procedure requires some degree of simplification of the phenomenon under study. However, no matter how sophisticated, all automated data extraction tools follow this essential underlying logic. Many have specialized search algorithms that allow the program to look for more fine-grained information of interest.

There are more than two dozen available automated content analysis tools. We briefly survey a few of them that we have used or are planning to use for our data collection efforts. AutoMap is an extraction tool that specializes in collecting information about key actors, their relationships, and their relation to an event or a set of events of interest. Automap also provides the attributes of actors including roles (leaders / followers), psychological factors, and resources. On the basis of such information extracted using AutoMap, we can then extract further information concerning groups and the entire structure of social networks of individuals and groups with the use of additional tools such as Organizational Risk Analyzer [Carley and Reminga, 2004]. T-Rex (The Resource Description Framework Extractor) uses cultural, economic, political, social, and religious variables provided by social scientists in conjunction with other data sources such as surveys and event databases and automatically extracts relevant data from news outlets, blogs, newsgroups, and wikis. STORY crawls the web at 50,000 pages/day and extracts facts and schemas. OASYS (Opinion Analysis System) is a specialized content analysis tool that is designed to extract information in real time from over 100 news sites in 8 languages and 12 countries regarding actors' opinions about any given topic, together with a measure of the intensity of these opinions. General Inquirer and ReadMe are more generic and less specialized examples of automated content analysis software that takes a set of text documents as input and processes these texts into various categories chosen by the user. The seminal tool of this field, KEDS (Kansas Event Data System), is specialized for generating event data and has the most extensive databases constructed using its system while Profiler+ performs leadership style analysis by looking for specific words that indicate leadership traits. Table 9.3 summarizes these tools by specialty.

The prospect of using these tools is exciting. However, there are at least six challenges of varying degrees of difficulty that confront potential users. As a test case, IndaSea helped us to use their CSM (Cultural Simulation Model) tool to profile President Musharraf of Pakistan. One strand of the results is shown in Table 9.4. Here we are looking at one of the standards of the GSP trees dealing with military doctrine – specifically, the tendency to shun violence.

- (1) Coverage, as already mentioned, is a concern with the databases, but not any more

Write out what
OASYS + STORY
stand for

TABLE 9.3 A Summary of Some Automated Data Extraction Tools

Name	Specialty
AutoMap	Relevant actor identification and relationship extraction
CSM by IndaSea	Generic
General Inquirer	Generic
KEDS	Event data extraction
OASYS	Opinion on any given topic
Profiler+	Leadership style/trait extraction
ReadMe	Generic
STORY	Generic
T-REX	Generic

TABLE 9.4 Example output of one of the automated web-scraping tools targeting a personality trait of an illustrative leader (Musharraf's tendency to shun violence)

<i>Standards: Shun Violence</i>	
Keywords	Peace (Some search terms for the opposite of shunning violence: military, army, tactics, strategy, operation, operations, tactic)
Positive Evidence	<p>05-06-2007 You are saying that this problem will not be solved by fighting, but it will not be solved by peace either. This problem will not be solved unless the main sources and centres of this movement – the Pakistani ISI – is not shut down.</p> <p>06-18-2007 In addition to disturbing the people in Afghanistan, the ISI has another programme which is to defame Islam across the world. It also comments on the Islamic instructions which urge people to make every possible effort to ensure peace and stability and prevent bloodshed.</p>
Negative Evidence	<p>04-01-2007 According to the Washington Post, officers within Pakistan's intelligence agency, the Inter-Services Intelligence Agency, proposed the following idea to address the vulnerability of its nuclear weapons to an Indian attack: "Let's hide them in Afghanistan-the Indians will never be able to attack them there."</p> <p>04-02-2007 There have been violent clashes in the tribal area in the past few days between tribal militia groups, which are difficult to distinguish from the Taliban linked to the ISI, and those groups that cooperate with foreign terrorists. In all likelihood, the ISI is behind this development, because an ISI patrol has been ambushed for the very first time, almost certainly in retaliation.</p> <p>04-10-2007 According to her, in the fall of 1993 her assassination was ordered and the "chosen assassin was a Pakistani with ties to the ISI during the Afghan Jihad. His name was Ramzi Yusef. He had participated in the first attack on the World Trade Centre in New York earlier that same year, on February 26."</p>

than it is with the newsfeeds. Any given country may or may not have an open, free press, so the viewpoints available and indeed, the veracity of what is published may be called into question. Where there is a free press, one must be sure that all views across the political spectrum are captured and appropriately tagged. These issues may render the newsfeed extraction problematic for certain of the parameter sets of interest. In Table 9.4, Musharraf's case, while there may not be a totally free press within Pakistan, the Pakistani president is a high visibility individual, and there is coverage at least by the Western press, a press with its own worldview.

(2) Another challenge in using automated content analysis tools lies in building the catalog that contains the necessary categories of key words and their combinations, both (or all) of which represent our model parameters. The main snag we face on this front lies in building a truly comprehensive and accurate catalog of keywords for the machine to use in extracting information from the exponentially growing quantity of available machine-readable text. Programming software that looks for such keywords and their combinations and counts their frequency is not difficult to build, and such programs already exist. There are also readily available generic catalogs or "dictionaries" that contain categories of words for us to import into a program and use for content analysis. However, these categories are sometimes too broad and generic for our purpose of extracting very specific information of interest, and thus these available "dictionaries" may be of only limited use. Table 9.5 shows a simple search with a single keyword ("peace").

(3) In addition to having proper keyword synonyms, it also may be that a schema or model of a given parameter has to be constructed to accommodate the interpretation and transformation of proxy variables. For example, in the row designated "positive evidence" in Table 9.4, the second item found is not necessarily proof that Musharraf (via the ISI (Inter-Services Intelligence)) shuns violence so much as it is evidence that he has a program to defame Islam. This suggests that perhaps a schema of typical human actions that do and do constitute "shunning violence" might be able to weed out such an item by classifying it as "inconclusive." It is possible that this can be learned automatically if we provide our prior hand-coded models. However, that is still an untested assumption.

(4) One needs to test the error rates of all the extraction tools. This implies assembling a test corpus in addition to a training data set where all the ground truth is known. One can then measure precision and recall rates and determine if the extraction tools are doing a credible job or not. While they may work well on other test sets for which they were designed for, one must always recalibrate their performance for the types of searches and extractions of interest to a given model. For example, Table 9.4 shows the retrieval of two bits of positive evidence and three bits that are negative. Are these all the items to be found? If we extend the keyword list and add a model/schema of the behavior, what happens to precision and recall? This is an aspect of our project that merits some research.

(5) Seamless integration is a desirable objective and one would like whatever extraction technology is adopted to be invisible to end users. In many cases there are setup issues and challenges for using the output of these tools. This is mostly a question of effort needed to embed these tools so the end user will not need to deal directly with them.

(6) Even if all the other challenges are eliminated, a remaining issue is how to weigh all the evidence collected assess its reliability, and transform it into actual parameter estimates. For example, how do we get the computer to summarize Table 9.4? Is it just a matter of adding two positives and three negatives? Obviously that would be simplistic and misleading, particularly when some of the positives are mild or inconclusive whereas some of the negatives are extreme items such as assassination attempts that may or may not be tied to Musharraf's ISI (Inter-Services Intelligence). How do we combine such bits of evidence? Some of the more difficult aspects of a text and knowledge extraction tool have

to do with understanding a personality and determining its underlying motivations. This is a hard problem, and human analysts who work at a “country desk” or a “leader desk” tackle it according to a well-developed methodology that we recap in Section 9.7.

9.6 Overview of Subject Matter Expert Studies/Surveys

The most obvious and intuitive method of obtaining information we need for our model parameters is to simply to ask subject matter experts (SMEs) to provide this information in our preferred format for our countries of interest. Let us suppose that we are modifying and populating our joint PMFserv/FactionSim framework to build a virtual country for the purpose of, say, better understanding and simulating potential political instabilities in this particular country of interest. In this scenario, we would be particularly interested in modeling and analyzing how Diplomatic, Intelligence, Military and Economic (DIME) actions might affect the Political, Military, Economic, Social, Informational, and Infrastructure (PMESII) systems of the country, and, given the importance of this kind of project, we would like to use the most up-to-date and accurate information for our country of interest. Knowing the limitations of the two previously discussed means of extracting information—namely, country databases and automated data extraction tools—in the short term at least, we might in fact be better off by gathering information directly from the best available country experts, tapping their expertise by means of a survey questionnaire to them or by conducting open-ended interviews. For our purposes, administering a structured, self-explanatory web survey tailored to elicit exactly the information we need would in most cases be preferable to conducting unstructured, open-ended interviews (partly because these interviews would elicit a wealth of information that would then need to be sorted and coded).

There are three main difficulties associated with using subject matter experts to elicit the information we need. First, administering interviews with experts in either form – expert survey or open – ended interview—requires significant financial and human resources. This method of collecting information costs at least as much as – and in most cases considerably more than – the other previously discussed options. Unless we are fortunate enough to have high quality SMEs available to us on a pro-bono basis, seeking their expertise for a task such as filling the more than fifty parameters for the GSP Tree alone may be prohibitively complicated and expensive. Second, subject matter experts, by definition of being subject matter experts and by virtue of being human and therefore fallible, may sometimes provide us with biased and, from time to time, even blatantly incorrect information: e.g., see [Tetlock, 2005; Heuer, 1999]. To limit this bias, we would probably want to consult more than one subject matter expert on any particular country or topic. More importantly, being a country expert does not mean that one has complete and comprehensive knowledge; a country expert does not know everything there is to know about a country. Third and finally, simply finding subject matter experts for a particular country of interest may by itself pose a significant challenge. Social scientists, historians, and area studies scholars with specific country expertise are not in short supply, but their expertise is not evenly distributed around the globe; given the structural constraints that exist in academia, certain parts of the world and certain countries receive disproportionate attention, while others are relatively neglected (for example, there is a glut of available expertise on China, but much less expertise on countries such as Bangladesh and Fiji). In sum, while at first this most direct route of getting parameters from experts looks easy and straightforward, it is also beset with difficulties.

Authoring a survey (or a set of slider-bar GUI screens) that is self-explanatory and has validated questions about each parameter needed in a socio-cognitive agent model is time

consuming, but not intellectually difficult. Such a survey is needed for eliciting knowledge from country or leader desk experts. A different approach that is an extension of this one involves a distributed set of experts, each knowledgeable about a sub-part of the ethno-political region to be modeled. The US military today, for example, currently plans for three sets of multi-person teams to perform this task for an area of operations. These three types of teams were listed earlier in Figure 9.1 as the Human Terrain Team (HTT), Intel and Every Soldier a Sensor (ESS) team, and Civil Military Affairs (CMA) team. The Human Terrain team includes anthropologists and social experts who collect data that is directly relevant to profiling agent personas, their clan structures, attributes, and kinship links. The Intel and Every Soldier a Sensor group tends to collect data pertaining to biometrics, demographics, intent, and information flow patterns in the target region. The Civil Military Affairs team focuses on quantitative estimates of resources, facilities, jobs, economic activity, infrastructure, and the like. For a model like FactionSim-PMFserv, all of this data is important. At present most of it is collected and entered into databases. In the future, one can envision the agent-based models as being the main repositories of such information. This would both improve the data collection focus and provide tools for the analyst and trainee that are sensitized to the DIME-PMESII issues of the area of operation. Getting to that point may be a grand challenge worthy of a DARPA (Defense Advanced Research Projects Agency) style program given the scale-up entailed by such a distributed activity.

9.7 Overview of Integrative Knowledge Engineering Process

Some of the more difficult aspects of a text and knowledge extraction tool have to do with understanding a personality and determining its underlying motivations. This is a hard problem, and human analysts who work at a “country desk” or a “leader desk” approach it according to a well-developed methodology, though even they are subject to errors of omission and commission, biases, or slipups. We have studied that methodology during the years of assembling the Athena’s Prism diplomatic role playing game [Silverman et al., 2005] and have adopted our own version of it for the leaders and followers we have profiled. We published an account of that methodology in [Silverman and Bharathy, 2005] and recap it very briefly here since it is the essence of the automated knowledge extraction workbench we are trying to assemble.

Since multiple sources of data are involved, a process is required to integrate and bring all the information together. We employ a process centered around differential diagnosis. This design is also based on the fact that directly usable numerical data are limited and one has to work with qualitative, empirical materials. Therefore, in the course of constructing these models, there is the risk of contamination by cognitive biases and human error.

The burden of this integrative modeling process is to systematically transform empirical evidence, tacit knowledge and expert knowledge from diverse sources into data for modeling; to reduce, if not eliminate, the human errors and cognitive biases (for example, for confirming evidence); to ensure that the uncertainties in the input parameters are addressed; and to verify and validate the model as a whole, and the knowledge base in particular.

For lack of a better term, the process has been conveniently referred to as a Knowledge Engineering (KE) process due to extensive involvement of knowledge engineering techniques and construction of the knowledge models. A diagrammatic representation of the knowledge

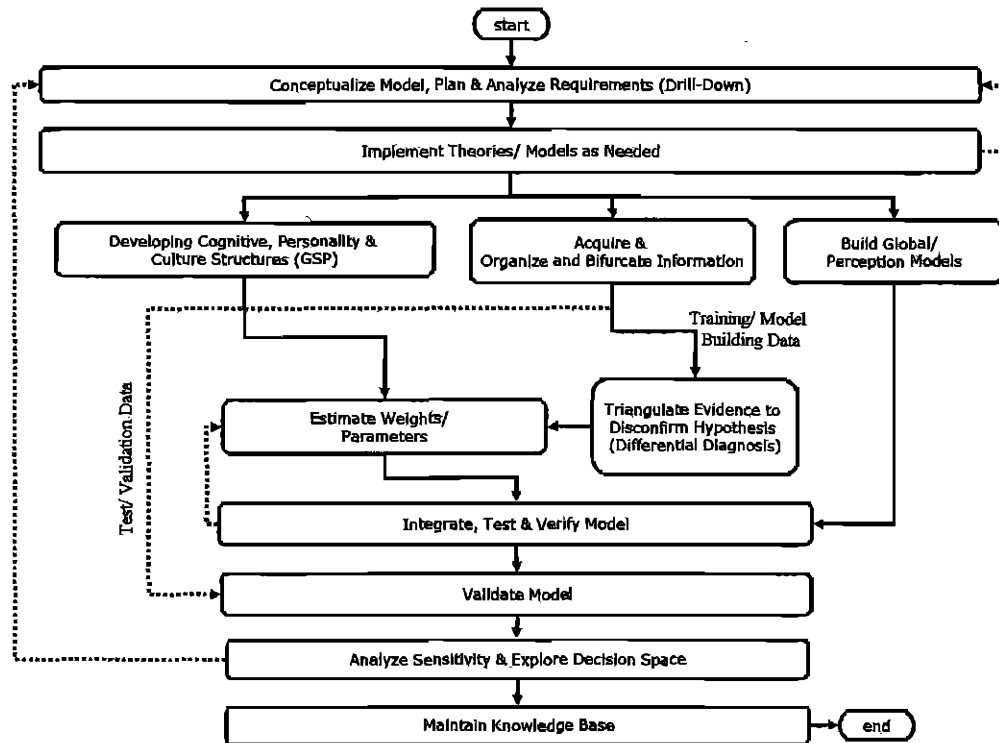


FIGURE 9.5 Knowledge Engineering Process Summary

engineering process is given in Figure 9.5.* The details of the process are beyond the scope of this chapter, but a summary of the methodology has been given below in Figure 9.5. Let us describe the salient features of the method.

Firstly, the body or corpus of qualitative information from different sources are aggregated and thematically organized in an evidence table. The input from experts and country database output, which directly pertains to the parameters, may be employed to help set the initial parameter values in the model, while anecdotal expert inputs and tangential estimates from the country databases are also incorporated in the evidence table. In order to ensure separation of model building (training and verification) and validation data, the empirical materials concerned are longitudinally divided into two different parts. One part is set aside for validation. The model is constructed and verified using the remaining part.

Since organizing information from otherwise diverse or amalgamated sources is critical to the success of the remaining modeling activities, a modified content analysis process is employed to collate and organize the evidence thematically. The themes of relevance are obtained from the high-level goals, standards, and preferences as well as from people and general potential behaviors of interest in the domain. The body of materials describing behavior in the model is split up into records, with each representing one and only one theme. Then, these are assigned theme codes, relevance, and reliability (subjective estimate of the source or info), and sorted according to the themes. The output is organized information

*This is a simplified view. Full details can be found in [Bharathy, 2006].

in tabular form with additional attributes such as reliability, frequency of occurrence, and relevance. Alternative hypotheses are selected at this juncture. The following table shows an excerpt from the evidence table, pertaining to the behavior of Richard, the Lionheart in the Crusade.

TABLE 9.5 Sample Evidence Table

Theme	Evidence	Reliability	Relevance
Economy	Amasses wealth in battles	Very High	...
Military	Conquers territory I, etc.	Very High	...

Having collected the data, one must integrate the data to arrive at the estimate of the parameters. Several tools and techniques have been devised for this purpose. Among them are tools for differential diagnosis (differential diagnoser) and pairwise comparison, which help elicit parameters in the graphical models through a systematic, defensible and transparent process. These tools accompany a mathematical framework, and contain provisions for estimating uncertainties in the expert inputs and empirical evidence.

When constructing models of behavior from evidence (be it from empirical evidence taken from the literature or from expert input), the modelers (or experts) employ the cognitive process of determining the motives of someone's behavior. This is subject to several biases [Kahneman et al., 1982; Gilovich and Griffin, 2002]. Foremost among these is confirmation bias, which may also subsume other biases such as availability bias and attribution bias. Simultaneous evaluation of multiple, competing hypotheses is very difficult to do and is against the cognitive bias of the human mind. Without an instrument designed to counter this fundamental bias, a modeler attempting to build Value Tree Models may easily be misled. The tendency to build models by confirming a plausible but favorite hypothesis will have enormous and grave implications at the next stage, when collating evidence provided by experts and empirical materials and building Value Tree Models. The human mind works though a "satisfying strategy." The process of selecting a favorite hypothesis is highly influenced by one's own conditioning, and the tendency is to see what one is looking for, and to overlook alternatives. Assessing evidence and attributing behavioral traits should ensure that no external cause explains the same behavior, that other competing traits do not explain the same behavior, and that confirmation bias is eliminated by a disconfirming hypothesis [Gorman and Gorman, 1984; Heuer, 1999].

In order to minimize the risk of not considering alternatives and considering non-diagnostic evidence, we have provided a tabular design, to carry out Differential Diagnosis, for generating and weighting alternative hypotheses, explanations or conclusions. Accordingly, the process forces one to look at competing hypotheses, and methodically disconfirm these alternatives rather than simply confirming a first hypothesis using available evidence. Once again, testing the usefulness and effectiveness of this differential diagnosis tool will be an important part of this project. Later, we will look at how this tabular structure can be exploited quantitatively.

The hypotheses in this case are parameters such as the nodes in the Value Trees (goals, standards and preferences of the characters being assessed). On the left, the framework includes the Organizer containing key evidence, thematically coded and attributed with reliability and relevance. However, the tool for Differential Diagnosis is centered on the hypotheses and evidence. Essentially, the hypotheses are pitted against the evidence through this matrix. If reliable evidence rejects a hypothesis, then the likelihood of that hypothesis is diminished significantly. We advocate including reliability and relevance for each piece of

evidence. Relevance (from the Organizer) identifies which items are most helpful in judging the relative likelihood of the hypotheses, and helps control the time spent on what appears to be irrelevant evidence. Using the Bayesian framework, we have developed a quantitative technique for the differential diagnosis, which attributes higher weight (an order of magnitude or more) to disconfirming evidence. We also established that otherwise rare events weigh more when they do occur as evidence. Essentially, the hypotheses are pitted against the evidence through this matrix.

The approach we suggest is to take all competing hypotheses (H_j) that explain a set of evidence and then pit them against this evidence (E_i). We find it best to work with a confirmation index that weighs disconfirming evidence about an order of magnitude higher than confirming evidence. Let us call this process of estimation based on disconfirming evidence "differential diagnosis," a term found in medical decision-making. Differential diagnosis is nothing but a triangulation technique. This technique much used in the field of medicine, where observing and discovering evidence leads gradually to a consideration of a short list of illnesses most likely to be behind a particular set of symptoms. While we share with medical diagnosis the same intent of unearthing the most likely hypotheses, there are some minor differences in terms of means and ends. Our purpose is not to identify a specific cause, but instead to attribute behavioral evidence to hypotheses of causes. While medical diagnosis tends to favor testing hypotheses largely serially and in a qualitative fashion, our models involve running several tests simultaneously by considering a set of competing hypotheses and triangulating a large set of evidence, as quantitatively (or quasi-quantitatively) as possible.

While both methods give more weight to disconfirming evidence, we give some (but lesser) consideration to confirming or supporting evidence. While disconfirmation is a much more powerful technique compared to confirmation, the latter provides some weak, yet economical, diagnosis in the absence of disconfirming evidence. We also take into consideration the reliability of data and typicality of events. The main difference might be in our employment of an explicit and simple tool that is amenable to both Bayesian analysis and also simple, score-based decision support.

This results in the following simplified expression for a metric called Confidence Index. Mathematically, Confidence Index (CI_{Avg}) for a given Hypothesis (H_j) may be defined as the weighted average measure of all the confirmations (and disconfirmations) associated with a hypothesis (with the subscript denoting that it is an average index over the given hypothesis):

$$CI_{Avg}(H_j) = \frac{1}{n} \times \sum_{i=1}^n K \times C_{ij} \times R_i \times \frac{f_i}{f_{R_i}}$$

where $K = \{w_1 \text{ when } C_{ij} \geq 0, \text{ and } w_2 \text{ when } C_{ij} < 0\}$.

Essentially, K is used to assign a higher weight (say an order of magnitude) to disconfirming evidence ($w_2 \gg w_1$). We have used w_1 value of 1 and w_2 value of 20. f_i is the frequency of the evidence, if the evidence given summarizes separate occurrences of behavioral evidence. Similarly, f_{R_i} is the typicality of the evidence (indicator of frequency of seeing that type of evidence in the real world. That is a measure of $P[E_i]$). Reliability (R_i) is subjectively estimated based on the source of the evidence as well as the confidence with which the evidence has been outlined by the source. For the sake of illustration, ignoring f_i and f_{R_i} , the expression for CI_{Avg} , this may also be simplified as:

$$\frac{1}{n} \times \sum_{i=1}^n K \times C_{ij} \times R_i$$

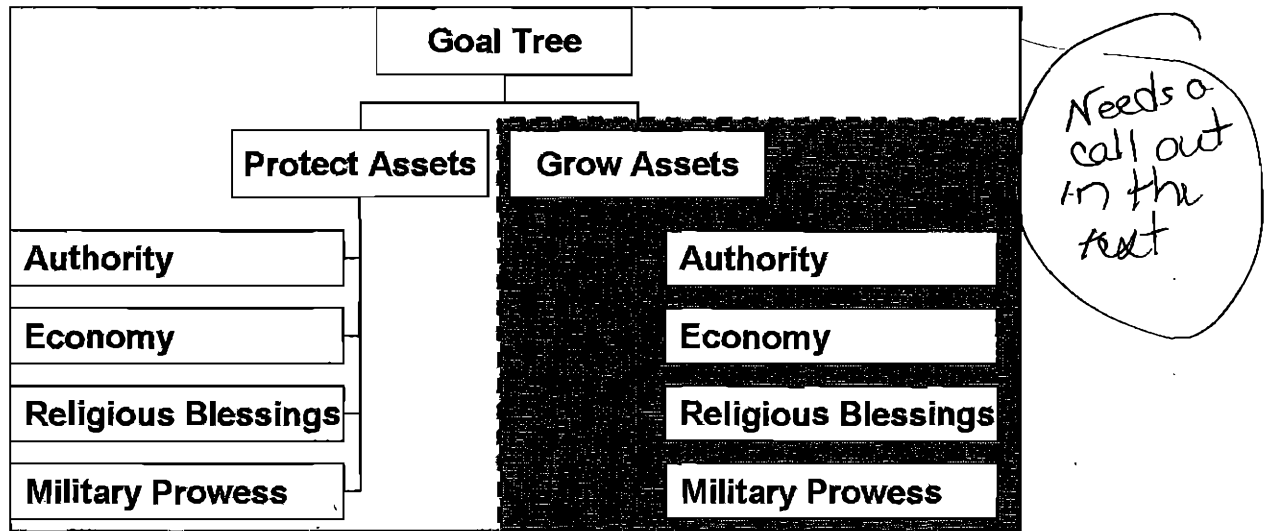


FIGURE 9.6 Segment of Goal Tree

The competing hypothesis that has the highest positive confidence wins only if the hypotheses are mutually exclusive, if the difference in CI is significant ($CI_{Avg} > 1.0$), and if the variance is small. For hypotheses which are not mutually exclusive, ordinal ranking might be obtained. When mutually exclusive hypotheses cannot be clearly distinguished by their confidence score, multiple competing hypotheses might have to be entertained during the course of the sensitivity analysis. Differential diagnosis allows one to consider all relevant evidence at once, and also gives higher weight to disconfirming evidence as described above. It allows one to find out whether these hypotheses could be ranked in the context of all available evidence. The details pertaining to the derivation and use of differential diagnosis have been taken from [Bharathy, 2006].

Now, let us consider the following cases to illustrate this technique. Differential diagnosis in the Crusade example has been illustrated through the following stylized cases. Note that simplifications have been made to introduce and illustrate the technique.

Example Question Again, consider the character of Richard, the Lion Heart. There are a few hypotheses (that could form the basis for some selected nodes of the GSP Tree) offered to explain Richard's spending time on a number of wars. Is Richard's inclination to grow any of the following resources (expansion of empire, wealth, religious blessings, or military prowess) more influential than other inclinations in explaining his behavior? Could his inclinations be ranked?

Let us formulate the above questions into the following competing hypotheses: Growth and expansion of the empire (Authority: H1), wealth (Economy: H2), religion (Religion: H3), or whether he loves warfare for its own sake (Military prowess: H4).

- H1: Grow Authority, Expand and Rule:** Richard is an expansionist and wanted to expand his kingdom and authority.
- H2: Protect Authority & Govern:** Richard fought to protect his authority from enemies, and the wars were thrust upon him.
- H3: Fill the Coiffers/ Max Econ Benefits:** Richard wants to grow his economic assets through fighting wars.

TABLE 9.6 Stylized Example of Differential Diagnosis

Stylized Example of Differential Diagnosis			Hypotheses						
			H1	H2	H3	H4	H5		
M1	Amasses wealth in battles	V. High			√				High
M2	Conquers territory 1, territory 2 etc.	V. High	√	√	√	√	√	×	Low
E2	Sells territory conquered	V. High	×	×	√				Medium
A1	Seldom governs the lands he has authority over.	V. High	×	×				×	Low
E1	Spends excessively on battles.	V. High	√		×		√	×	High
E3/ A3	Known to have said: "If I could find a buyer, I would mortgage London to raise money for battles"	Medium	×	×	√		√		Medium
A4	Spends most of his 9 year rein outside England on Crusade	V. High	√	×		√	√	√	Low
R1	Has not obeyed other religious laws	Moderate				×		×	High
E2	Fought against his father first	V. High	√			×	√	×	Low
E4	Died fighting over a treasure	V. High			√				Low
Confidence Index									

H4: Religious Duty: Richard wants to protect and grow his religious blessings.

H5: War for the War's Sake: loves the battlefields and wants to fight wars for the wars sake.

Then, we construct the table (Tables 9.6, 9.7)* and pit all of these hypotheses against the available sets of evidences. As one can see in Tables 9.6, 9.7, a number of rows of evidence disconfirm Richard's religious inclination, while there is little that contradicts Richard's inclination to grow military assets. It should not surprise the reader that Richard seems most inclined to grow his military prowess, followed by his desire for wealth, and his desire to govern, in that order. Therefore, this is a behavior that may not provide much additional information for identifying and sifting through his values. However, his other behaviors begin to contradict some of the existing the hypotheses.

In addition to the above use of differential diagnosis, where we illustrated the process of disconfirming hypothesis with available evidence, the same technique could be employed in different forms. For example, an expert could be encouraged to come up with different plausible scenarios. Once such a set of scenarios has been gathered and recorded, the expert could be asked to carry out differential diagnosis using these scenarios. The expert then attempts to disconfirm the hypothesis using the scenarios he or she has generated. This thought experiment could work as a powerful technique.

Introspection, Revision and Dialog: The degree of disagreement can be used to generate feedback to the experts themselves, and their assumptions can thereby be made transparent. Then the expert can redesign the GSP Tree while con-

*For the sake of simplicity, I have used the expression that $CI_{Avg} = 1/n \sum_{i=1}^n K \times C_{ij} \times R_i$, where $K = w_1$ when $C_{ij} \geq 0$ and $K = w_2$ with $C_{ij} < 0$. Essentially, K is used to assign a higher weight (say an order of magnitude) to disconfirming evidence ($w_2 \gg w_1$). Other forms of this relationship are being investigated. Additionally, thought experimenting the plausibility of generated scenarios disconfirming hypotheses may be employed as another input in this process.

TABLE 9.7 Stylized Example of Differential Diagnosis

Stylized Example of Differential Diagnosis									
M1	Amasses wealth in battles	0.9			+0.3				High
M2	Conquers territory 1, territory 2 etc.	0.9	+0.9	+0.6	+0.3	+0.8	+0.9	-0.8	Low
E2	Sells territory conquered	0.9	-0.9		+0.9				Medium
A1	Seldom governs the lands he has authority over.	0.9	-1.0	-1.0				-0.5	Low
E1	Spends excessively on battles.	0.9	+0.5		-0.9		+0.8	-0.8	High
E3/A3	Known to have said: "If I could find a buyer, I would mortgage London to raise money for battles"	0.5	-0.9	-1.0	+0.5		+0.9		Medium
A4	Spends most of his nine-year reign outside England on Crusade	0.9	+0.7			+0.7	+0.8	+0.5	Low
R1	Has not obeyed other religious laws	0.5				-0.6		-0.5	High
R2	Fought against his father first	0.9	+0.5				+0.8	-0.8	Low
E4	Died fighting over a treasure	0.9			+0.9	-0.2			High
	Confidence Index (with K=1 when Cij>=0, K=20 when Cij<0)		-0.9	-2.8	-1.8	-0.9	-0.5	-1.7	

sciously bracketing one or more assumptions. This kind of exercise can also be useful in group sessions to discuss the differences. In essence, it can create introspection and dialog, which will often focus attention on the root of the actual problem being studied. In a more superficial treatment, a structure could be adopted through a consensus seeking process, or by bootstrapping, or differently weighing expert and lay designs.

Uncertainty Estimation: The estimates can also provide estimates of the uncertainty (or confidence) in the GSP Tree.

We have employed this process manually in the past to create several models of leaders, followers, crowd members, rebels, agitators in conflict situations. We have been able to validate our integrative process under naturalistic conditions by testing, verifying and validating these models. As mentioned earlier, the process does get very laborious when constructing multiple models by hand. Therefore, we are in the process of automating the previously described manual process, incorporating text-mining, semantic analysis as well as Bayesian update.

9.8 Concluding Remarks

Our community would be remiss if it did not try to respond to the ideas of leaders in military and diplomatic circles who are now facing the challenge of promoting deeper thought, creating rehearsal environments, and developing analytic capability about cultural issues and local population needs/wants around the globe. They have funded programs that collect country data and conduct link analysis and social network studies. At the same time, they may lack the experience or expertise to appreciate the tools that the field of human behavior modeling currently has to offer, or is now in the process of developing.

In this chapter, we have argued that the available country datasets are an invaluable

resource that will permit us in the human behavior M&S (Modeling and Simulation) field to more realistically profile factions, and their leaders and followers. This in turn will help us to develop tools for those interested in analyzing alternative competing hypotheses for DIME-PMESII (Diplomatic, Informational, Military, and Economic actions – Political, Military, Economic, Social, Informational, and Infrastructure effects) studies. At the same time, there are significant growing pains and challenges involved in trying to put the country data to use. This chapter reviewed those challenges by looking at three pathways for extracting and parameterizing the data – webscraping of newsfeeds, extracting and translating data from country databases, and (semi-) automated surveying (i.e., web questionnaires with data translation and model instantiation capacity) of subject matter experts. In each of these areas there are significant challenges and obstacles to seamless integration, not the least of which is that profiling individuals and groups is difficult even for the smartest humans. By using a triangulation of the three approaches, and a knowledge engineering approach that mimics how country and leader experts currently do the job (alternative competing hypotheses), we believe that one can move ahead as outlined in this chapter.

This chapter examined how to use this approach with the help of a case study involving a socio-cognitive agent architecture (FactionSim-PMFserv). The hope is that the automated extraction will speed the development of gameworlds and scenarios with a tool like this. This push seems doubly pertinent since a parallel development in recent years has been the scientific struggles of those working to unify multi-resolution frameworks that permit modeling “deep” modeling of a small number of cognitively-detailed agents able to interact with and influence large numbers of “light” socio-political agents. This work is necessary if we are to have more realistic “socio-cognitive” agents, ones that are useful for the types of analysis and training/rehearsal M&S worlds envisioned here. This is part of the wider effort to have more realistic agents and detailed worlds that influence their decisions.

The validity of the models and theories inside the agents has not been a focus of this chapter. However, “correctness” is in equal parts about the data used and the generative mechanisms inside the agents. Both of these are finally more important than whether any particular predictions turn out to be accurate. Much of this chapter dealt with how to obtain the best possible data. We should close by also pointing out that if the generative mechanisms are roughly or in principle “correct,” then one can trust that experiments with the agents will yield useful insights about various policies and how these policies in turn will influence the agents. That is why one attempts to equip social agents with more and more advanced cognitive capabilities. This work suggests some words of advice and also caution to those attempting simulations with various country databases – start with best available models (with higher internal validity), then conduct adequacy tests, validity assessments, and replication of results across samples. Even after all that, social system simulations will rarely yield precise forecasts and predictions. Rather, their utility lies in exploring the possibility space and in understanding mechanisms and causalities so that one can see how alternative DIME (Diplomatic, Informational, Military, and Economic) actions might lead to the same or unexpected PMESII (Political, Military, Economic, Social, Informational, and Infrastructure) effects.

Acknowledgement

This research was partially supported by AFOSR, DARPA, and the Beck Fund, though no one except the authors is responsible for any statements or errors in this manuscript.

References

- R. Axelrod. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, NJ, 1998.
- G. K. Bharathy. *Agent Based Human Behavior Modeling: A Knowledge Engineering Based Systems Methodology for Integrating Social Science Frameworks for Modeling Agents with Cognition, Personality and Culture*. PhD thesis, University of Pennsylvania, 2006.
- K. M. Carley and J. Reminga. ORA: Organization risk analyzer. Technical Report CMU-ISRI-04-106, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, 2004.
- K. M. Carley, J. Diesner, and M. D. Reno. AutoMap user's guide. Technical Report CMU-ISRI-06-114, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, 2006.
- C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff, and V. S. Subrahmanian. OASYS: An opinion analysis system. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, pages 21–26, 2006.
- C. Cioffi-Revilla and S. P. O'Brien. Computational analysis in US foreign and defense policy. Paper prepared for the *First International Conference on Computational Cultural Dynamics*, University of Maryland, College Park, MD, August 2007.
- Comparative Study of Electoral Systems Secretariat. Comparative study of electoral systems 2001-2006. Computer File, 2004. ICPSR version. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer] 2003. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].
- J. M. Epstein. Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences*, 99(3):7243–7250, 2002.
- D. C. Esty, J. A. Goldstone, T. R. Gurr, B. Harff, P. T. Surko, A. N. Unger, and R. Chen. The state failure project: Early warning research for US foreign policy planning. In J. L. Davies and T. R. Gurr, editors, *Preventive Measures: Building Risk Assessment and Crisis Early Warning Systems*, chapter 3. Rowman and Littlefield, Boulder, CO and Totowa, NY, 1998.
- European Values Study Group and World Values Survey Association. European and world values surveys four-wave integrated data file, 1981-2004, 2006. v.20060423.
- M. Fayzullin, V. S. Subrahmanian, M. Albanese, C. Cesarano, and A. Picariello. Story creation from heterogeneous data sources. *Multimedia Tools and Applications*, 33(3):351–377, 2007.
- T. Gilovich and D. W. Griffin. Heuristic biases: Then and now. In T. Gilovich, D. W. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 1–18. Cambridge University Press, Cambridge, UK, 2002.
- Global Barometer Surveys Program. Global barometer surveys. <http://www.globalbarometer.net/index.htm>. Accessed November 2007.
- M. E. Gorman and M. E. Gorman. A comparison of disconfirmatory, confirmatory and control strategies on wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology Section A*, 36(4):629–648, 1984.
- R. F. Harrod. A second essay in dynamic theory. *Economic Journal*, 70:277–293, 1960.
- Heidelberg Conflict Barometer. Heidelberg conflict barometer. <http://www.hilk.de/en/konflikbarometer/index.html>. Accessed May 2008.
- M. G. Hermann. Who becomes a political leader? Lead-

spacing

- ership succession, generational change, and foreign policy. http://www.allacademic.com/meta/p69844_index.html, a. Accessed November 2007.
- M. G. Hermann. Assessing leadership style: A trait analysis. <http://socialscienceautomation.com/>, b. Accessed November 2007.
- R. J. Heuer. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, Washington, DC, 1999.
- A. O. Hirschman. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations and States*. Harvard University Press, Cambridge, MA, 1970.
- D. Hopkins, G. King, M. Knowles, and S. Melendez. ReadMe: Software for automated content analysis. <http://gking.harvard.edu/readme/docs/readme.pdf>, 2007.
- R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, and W. Gupta, editors. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Sage Publications, Thousand Oaks, CA, 2004.
- M. Johns. *Deception and Trust in Complex Semi-Competitive Environments*. PhD thesis, University of Pennsylvania, 2006.
- D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 1982.
- G. J. Kim and G. K. Bharathy. Penn conflict database catalogue. <http://www.seas.upenn.edu/~jiyunkim/pcdc.pdf>. Accessed November 2007.
- J. Kipp, L. Grau, K. Prinslow, and D. Smith. The human terrain system: A CORDS for the 21st century. *Military Review*, 86(5):8-15, September-October 2006.
- G. Lafree and L. Dugan. Global terrorism database, 1970-1977. Computer File, 2007. College Park, MD: University of Maryland [producer], 2006. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 04-04.
- W. A. Lewis. Economic development with unlimited supplies of labour. *Manchester School*, 22:139-191, 1954.
- I. S. Lustick, D. Miodownik, and R. J. Eidelson. Secessionism in multicultural states: Does sharing power prevent or encourage it? *American Political Science Review*, 98(2):209-229, 2004.
- N. Machiavelli. *The Chief Works and Others*. Duke University Press, Durham, 1965.
- N. Machiavelli. *The Prince*. Cambridge University Press, Cambridge, UK, 1988.
- S. Milgram. The small world problem. *Psychology Today*, pages 60-67, May 1967.
- Minorities at Risk Project. Minorities at risk project. <http://www.cidcm.umd.edu/mar/>. Accessed November 2007.
- J. Park and W. Fables. Perception based synthetic vision system. In *International Conference on Complex Systems*, Boston, MA, October-November 2007.
- R. W. Pew and A. S. Mavor, editors. *Modeling Human and Organizational Behavior: Application to Military Simulations*. National Academy Press, Washington, DC, 1998.
- Political Instability Task Force. Political instability task force worldwide atrocities dataset. <http://web.ku.edu/keds/data.dir/atrocities.html>. Accessed November 2007.
- P. M. Regan. *Civil Wars and Foreign Powers: Interventions and Intrastate Conflict*. University of Michigan Press, second edition, 2000.
- F. E. Ritter, M. N. Avraamides, and I. G. Councill. Validating changes to a cognitive architecture to more accurately model the effects of two example behavior moderators. In *Proceedings of the Eleventh Conference on Behavior Representation in Modeling and Simulation, SISO*, Orlando, 2003.
- M. R. Sarkees. The correlates of war data on war: An update to 1997. *Conflict*

spacing

- Management and Peace Science*, 18(1):123–144, 2000.
- F. Schneider and D. H. Enste. Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38:77–114, 2000.
- P. A. Schrodtt and D. J. Gerner. The kansas event data system: A beginner's guide illustrated with a study of media fatigue in the palestinian intifada. <http://web.ku.edu/keds/home.dir/guide.html>. Accessed 1996.
- B. G. Silverman and G. K. Bharathy. Modeling the personality & cognition of leaders. In *Proceedings of the Thirteenth Conference on Behavioral Representations in Modeling and Simulations, SISO*, May 2005.
- B. G. Silverman, M. Johns, R. Weaver, K. O'Brien, and R. Silverman. Human behavior models for game-theoretic agents. *Cognitive Science Quarterly*, 2(3/4):273–301, 2002.
- B. G. Silverman, R. L. Rees, J. A. Toth, J. G. Cornwell, K. O'Brien, M. Johns, and M. Caplan. Athena's prism – a diplomatic strategy role playing simulation for generating ideas and exploring alternatives. In *Proceedings of the First International Conference on Intelligence Analysis*, MacLean, VA, 2005.
- B. G. Silverman, G. K. Bharathy, K. O'Brien, and J. Cornwell. Human behavior models for agents in simulators and games: Part II: Gamebot engineering with PMFserv. *Presence: Teleoperators and Virtual Environments*, 15(2):163–185, 2006a.
- B. G. Silverman, M. Johns, J. G. Cornwell, and K. O'Brien. Human behavior models for agents in simulators and games: Part I: Enabling science with PMFserv. *Presence: Teleoperators and Virtual Environments*, 15(2):139–162, 2006b.
- B. G. Silverman, G. Bharathy, B. Nye, and R. J. Eidelson. Modeling factions for “effects based operations”: Part I – leaders and followers. *Computational & Mathematical Organization Theory*, 13(4):379–406, 2007.
- B. G. Silverman, G. K. Bharathy, B. Nye, and T. Smith. Modeling factions for “effects based operations”: Part II – behavioral game theory. *Computational & Mathematical Organization Theory*, 14(2):120–155, 2008.
- P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. *General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA, 1966.
- V. S. Subrahmanian. Cultural modeling in real time. *Science*, 317(5844):1509–1510, 2007.
- R. Sun. *Cognition and Multi-Agent Interaction*. Cambridge University Press, 2006.
- P. E. Tetlock. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, Princeton, NJ, 2005.
- University of Kansas. European protest and coercion data. <http://web.ku.edu/ronfran/data/index.html>. Accessed November 2007.
- Uppsala Conflict Data Program. Uppsala conflict database. <http://www.pcr.uu.se/database/>. Accessed November 2007.
- A. Varshney and S. Wilkinson. Varshney-wilkinson dataset on hindo-muslim violence in india, 1950–1995. Computer File, 2006. Cambridge, MA: Harvard University, WCFLA [producer], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] 02-17.
- M. D. Young. Building WorldView(s) with Profiler+. In M. D. West, editor, *Applications of Computer Content Analysis*, volume 17 of *Progress in Communication Sciences*. Ablex Publishing, 2001.

Spacing

city

1